

Can Turnover Go to Zero?

Zura Kakushadze^{§†1}

[§] *Quantigic[®] Solutions LLC*

*1127 High Ridge Road #135, Stamford, CT 06905*²

[†] *Department of Physics, University of Connecticut*

1 University Place, Stamford, CT 06901

(May 30, 2014; revised: July 14, 2014)

Abstract

Internal crossing of trades between multiple alpha streams results in portfolio turnover reduction. Turnover reduction can be modeled using the correlation structure of the alpha streams. As more and more alphas are added, generally turnover reduces. In this note we use a factor model approach to address the question of whether the turnover goes to zero or a finite limit as the number of alphas N goes to infinity. We argue that the limiting turnover value is determined by the number of alpha clusters F , not the number of alphas N . This limiting value behaves according to the “power law” $\sim F^{-3/2}$. So, to achieve zero limiting turnover, the number of alpha clusters must go to infinity along with the number of alphas. We further argue on general grounds that, if the number of underlying tradable instruments is finite, then the turnover cannot go to zero, which implies that the number of alpha clusters also appears to be finite.

Keywords: hedge fund, alpha stream, crossing trades, portfolio turnover, factor model, correlation structure

¹ Email: zura@quantigic.com

² DISCLAIMER: This address is used by the corresponding author for no purpose other than to indicate his professional affiliation as is customary in publications. In particular, the contents of this paper are not intended as an investment, legal, tax or any other such advice, and in no way represent views of Quantigic Solutions LLC, the website www.quantigic.com or any of their other affiliates.

1 Introduction and Summary

When multiple alpha streams are traded on the same hedge fund platform, if the execution platform allows this, it makes sense to cross trades between different alpha streams thereby saving on transaction costs.³ When internal crossing is employed, portfolio turnover is reduced.

As more and more alphas are added, generally the percentage of the dollar turnover with respect to the total dollar investment – which percentage we refer to simply as “turnover” – is expected to decrease. The more correlated the trades are, the more correlated the alphas are, and the more correlated the trades are, the lower the internal crossing is expected to be. Therefore, while turnover reduction is not necessarily a simple (*e.g.*, linear) function of alpha correlations, it is clear that it is somehow related to them, so one can try to model turnover reduction based on alpha correlations. In (Kakushadze, 2014a) a spectral model of turnover reduction based on the alpha correlation structure was proposed using *principal component* analysis of the alpha correlation matrix. A simplified version of this model was used in (Kakushadze and Liew, 2014) to argue that when the number of alphas is large, the portfolio turnover has a non-vanishing limit.

In this note we use a factor model approach – as *complementary* to the principal component approach of (Kakushadze, 2014a) – to address the question of whether the turnover goes to zero or a finite limit as the number of alphas N goes to infinity. We argue that the limiting turnover value is determined by the number of alpha clusters F , not the number of alphas N . This limiting value behaves according to the “power law” $\sim F^{-3/2}$. So, to achieve zero limiting turnover, the number of alpha clusters must go to infinity along with the number of alphas. We further argue on general grounds that, if the number of underlying tradable instruments is finite, then the turnover cannot go to zero, which implies that the number of alpha clusters also appears to be finite. The large N limit (Kakushadze and Liew, 2014), (Kakushadze, 2014a) plays an important simplifying role in our discussion.

The remainder of this paper is organized as follows. Definitions are in Section 2. We discuss a factor model approach for alpha streams in Section 3. In Section 4 we briefly review the spectral model of turnover reduction of (Kakushadze, 2014a). In Section 5 we use the factor model approach of Section 3 to study what happens to the turnover in the large N limit, including the effects of the alpha clusters, specific risk, off-diagonal elements of the factor covariance matrix, and style and general non-binary risk factors on the turnover. In Section 6 we discuss a practical application of our results in Section 5. In Section 7 we discuss methods for estimating lower

³ An illustrative discussion of internal crossing and its benefits can be found in (Kakushadze and Liew, 2014). A spectral model of turnover reduction, which we discuss in this note, was recently proposed in (Kakushadze, 2014a). For a partial list of hedge fund literature, see, *e.g.*, (Ackerman *et al*, 1999), (Agarwal and Naik, 2000a, 2000b), (Amin and Kat, 2003), (Asness *et al*, 2001), (Brooks and Kat, 2002), (Brown *et al*, 1999), (Chan *et al*, 2006), (Edwards and Caglayan, 2001), (Edwards and Liew, 1999a, 1999b), (Fung and Hsieh, 1999, 2000, 2001), (Kao, 2002), (Liang, 1999, 2000, 2001), (Lo, 2001), (Schneeweis *et al*, 1996).

and upper bounds for the number of alpha clusters. In Section 8 we further argue – this is a non-rigorous “theorem” of sorts – on general grounds that, if the number of underlying tradable instruments is finite, then the turnover cannot go to zero, which implies that the number of alpha clusters also appears to be finite. Our main result is given by Eqns. (29) and (83), and the “theorem” in Section 8.

2 Definitions

We have N alphas α_i , $i = 1, \dots, N$. Each alpha is actually a time series $\alpha_i(t_s)$, $s = 0, 1, \dots, M$, where t_0 is the most recent time. Below α_i refers to $\alpha_i(t_0)$.

Let C_{ij} be the covariance matrix of the N time series $\alpha_i(t_s)$. Let Ψ_{ij} be the corresponding correlation matrix, *i.e.*,

$$C_{ij} = \sigma_i \sigma_j \Psi_{ij} \quad (1)$$

where $\Psi_{ii} = 1$.

In the papers (Kakushadze and Liew, 2014) and (Kakushadze, 2014a) it was argued that in the large N limit the turnover does not necessarily go to zero and can go to a finite limit depending on the structure of the correlation matrix.⁴ Our goal in this paper is to address the following question: What would make the turnover go to zero in the large N limit?

3 Factor Model

Generally, the covariance matrix C_{ij} can have the following undesirable properties. First, it can be (nearly) degenerate. Second, it may not be positive (semi-)definite.

Near degeneracy is caused by alphas that are almost 100% correlated or anti-correlated and can be cured by simply removing such “redundant” alphas.⁵ However, in practice, near degeneracy is usually caused by the fact that $M < N$ (in fact, in most practical applications $M \ll N$) and only M eigenvalues of C_{ij} are non-zero, while the remainder have “small” values, which can be positive or negative. These small values are zeros distorted by computational rounding.⁶ In such cases, the solution is not to remove any alphas (as they are not necessarily “redundant”), but to deform the covariance matrix so it is positive-definite. One such deformation

⁴ Caveats and limitations of modeling internal crossing and turnover reduction using the correlation matrix were discussed in (Kakushadze and Liew, 2014) and (Kakushadze, 2014a), so we will not repeat them here.

⁵ See (Kakushadze, 2014a) for a more detailed discussion.

⁶ Actually, this assumes that there are no N/As in any of the alpha time series. If some or all alpha time series contain N/As in non-uniform manner and the correlation matrix is computed by omitting such pair-wise N/As, then the resulting correlation matrix may have negative eigenvalues that are not “small” in the sense used above, *i.e.*, they are not zeros distorted by computational rounding. The deformation method mentioned below can be applied in this case as well.

based on (Rebonato and Jäckel, 1999) was discussed in (Kakushadze, 2014a) (see Subsection 3.1 thereof). Another, perhaps more commonly used method,⁷ is to employ a factor model approach.

The aforementioned issue arising when $M < N$ conceptually is the same as the problem of modeling risk of a portfolio consisting of a large number N_S of stocks.⁸ Unless the number $M_S + 1$ of observations in the corresponding time series is large compared with the number of stocks, the covariance matrix based on such time series is not expected to be very stable. Furthermore, if $M_S < N_S$, as above, the correlation matrix is degenerate. One way to circumvent this problem in the case of stocks is that one builds a factor model, where instead of N_S stocks one deals with F_S risk factors, where $F_S \ll N_S$ and $F_S \lesssim M_S$.

The same can be done with alphas. Instead of N alphas, one deals with $F \ll N$ risk factors and the covariance matrix C_{ij} is replaced by Γ_{ij} given by

$$\Gamma \equiv \Xi + \Omega \Phi \Omega^T \quad (2)$$

$$\Xi_{ij} \equiv \xi_i^2 \delta_{ij} \quad (3)$$

where ξ_i is the specific risk for each α_i ; Ω_{iA} is an $N \times F$ factor loadings matrix; and Φ_{AB} is the factor covariance matrix, $A, B = 1, \dots, F$. *I.e.*, the random processes Υ_i corresponding to N alphas are modeled via N random processes z_i (corresponding to specific risk) together with F random processes f_A (corresponding to factor risk):

$$\Upsilon_i = z_i + \sum_{A=1}^F \Omega_{iA} f_A \quad (4)$$

$$\langle z_i, z_j \rangle = \Xi_{ij} \quad (5)$$

$$\langle z_i, f_A \rangle = 0 \quad (6)$$

$$\langle f_A, f_B \rangle = \Phi_{AB} \quad (7)$$

$$\langle \Upsilon_i, \Upsilon_j \rangle = \Gamma_{ij} \quad (8)$$

Instead of an $N \times N$ covariance matrix C_{ij} we now have an $F \times F$ covariance matrix Φ_{AB} , which is much more stable than C_{ij} .

One approach to constructing a factor model for alphas is to have F_{style} style risk factors and $F_{cluster}$ cluster risk factors. In the case of stocks, cluster risk factors are usually referred to as industry risk factors. Since here we are dealing with alphas, we will refer to such risk factors as cluster risk factors. Generally, clusters can be thought of as groupings of alphas based on some similarity criteria, which to a large extent boil down to how closely alphas are correlated with each other, albeit such similarity criteria need not be (but can be) based on sample correlation matrix – see below. On the other hand, style risk factors are not based on any similarity criteria; instead, they are based on some estimated (or measured) properties of alphas – thus,

⁷ More commonly used in the case of stocks, that is.

⁸ The high-level math formalism is the same, albeit details are different (Kakushadze, 2014b).

in a given cluster we can have alphas with vastly different values of a given style factor. In the case of alphas, the following style factors *a priori* appear possible, at least when the underlying tradables are stocks: 1) volatility, 2) turnover, and 3) momentum. Another (perhaps more difficult to implement) style factor one may wish to consider is capacity, *i.e.*, how much capital each alpha can absorb. One may also wish to add other style factors depending on how alphas are constructed, *etc.*⁹

In the case of stocks, cluster factors are (usually) based on industry classification. In the case of alphas, one can use a taxonomy of alphas, *i.e.*, one classifies alphas according to how they are constructed – if the required data is available, that is. Out of thousands of alphas one may construct, many are very similar to each other by construction. It is then clear that this similarity makes them more correlated, just as stocks belonging to the same industry are more correlated.¹⁰ Just as in the case of stocks, it therefore makes sense to treat clusters as risk factors and model correlations between alphas based on such risk factors as opposed to computing them directly based on a large number N of the time series corresponding to individual alphas. Note that a bonus of the factor model approach is that the matrix Γ_{ij} is automatically positive-definite if all ξ_i are non-zero and the factor covariance matrix is positive-definite (albeit this condition can be relaxed).

4 Spectral Model

In the paper (Kakushadze, 2014a) we discussed a spectral model of turnover reduction, which suggests that the behavior of the turnover in the large N limit can be approximated as follows:¹¹

$$T \approx \rho_* \sum_{i=1}^N \tau_i |w_i| \quad (9)$$

where (the modulus $|\cdot|$ stands for the absolute value of the sum)

$$\rho_* \equiv \frac{\psi^{(1)}}{N\sqrt{N}} \left| \sum_{i=1}^N V_i^{(1)} \right| \quad (10)$$

⁹ For a discussion of a general framework for alpha factor models, see (Kakushadze, 2014b).

¹⁰ Conversely, if it were the case that $M \gg N$, high (anti-)correlations of alphas in the sample correlation matrix would imply their similarity. However, in practice we have $M < N$ and, in fact, $M \ll N$, so deducing similarity from the sample correlation matrix is not as straightforward. There exist (typically, proprietary) algorithms for building binary clusters based on the sample correlation matrix and/or its principal components. Such clusters generally are expected to inherit, at least to a large extent, the usual out-of-sample instability of the off-diagonal elements of the sample correlation matrix. However, in cases where the detailed information about how alphas are constructed is not available – *e.g.*, the only data available could be the position data, *i.e.*, vectors of desired holdings to achieve by some times T_1, T_2, \dots – then such algorithms might come handy.

¹¹ Let us note that, in the paper (Kakushadze, 2014a), Eq. (10) below was obtained by observing that in the context of the spectral model the largest eigenvalue of Ψ_{ij} has the leading contribution to T in the large N limit. If other contributions are not suppressed, then (10) can be thought of as an approximate lower bound for turnover reduction.

and $\psi^{(1)}$ is the largest eigenvalue of the correlation matrix Ψ_{ij} , while $V^{(1)}$ is the corresponding right eigenvector of Ψ_{ij} normalized such that¹²

$$\sum_{i=1}^N \left(V_i^{(1)}\right)^2 = 1 \quad (11)$$

Also, w_i are the weights with which alphas are combined, $\sum_{i=1}^N |w_i| = 1$, and τ_i are the turnovers corresponding to individual alphas α_i .

For illustrative purposes, let us consider the case where all off-diagonal elements of the correlation matrix Ψ_{ij} are identical: $\Psi_{ij} = \rho$ ($i \neq j$). Also, let all $T_i \equiv \tau_i |w_i|$ be identical, let $\psi^{(p)}$ be the eigenvalues of Ψ_{ij} , and let $V^{(p)}$ be the corresponding eigenvectors. Then, in the basis where $\psi^{(1)}$ is the largest eigenvalue, we have (note that $V_i^{(1)} \equiv 1/\sqrt{N}$, $i = 1, \dots, N$ in this case):

$$\sum_{i=1}^N V_i^{(1)} T_i = \frac{1}{\sqrt{N}} \sum_{i=1}^N T_i \quad (12)$$

$$\sum_{i=1}^N V_i^{(p)} T_i = 0, \quad p > 1 \quad (13)$$

$$\psi^{(1)} = 1 + (N - 1) \rho \quad (14)$$

$$\psi^{(p)} = 1 - \rho, \quad p > 1 \quad (15)$$

$$\rho_* = \frac{1 + (N - 1) \rho}{N} \quad (16)$$

which reproduces Eq. (12) in (Kakushadze and Liew, 2014). The first two equations above capture the essence of why the spectral model of (Kakushadze, 2014a) is expected to be a good approximation in the large N limit even in the case of a general correlation matrix: for generic configurations of T_i (*i.e.*, such that T_i are not highly skewed and are reasonably distributed around their mean value), the contribution to ρ_* due to the first principal component (corresponding to the largest eigenvalue) is dominant,¹³ while contributions due to other principal components are suppressed by powers of $1/N$. In the above case of uniform correlations, let us consider two extreme values of ρ . First, when $\rho = 1$, *i.e.*, when all alphas are 100% correlated with each other, we have $\psi^{(1)} = N$ and $\rho_* = 1$, and there is no turnover reduction. On the other hand, when $\rho = 0$, *i.e.*, when none of the alphas are correlated with each other, we have $\psi^{(1)} = 1$ and $\rho_* = 1/N$, and the turnover goes to zero in the large N limit. The above extreme cases are in agreement with what we expect based on an intuitive picture of turnover reduction (Kakushadze and Liew, 2014).

¹² Here the basis of α_i is taken (*i.e.*, the signs of α_i are chosen) such that $\sum_{i,j=1}^N \Psi_{ij} \equiv N^2 \rho' = N(1 + (N - 1)\bar{\rho})$ is maximized ($\bar{\rho}$ is the mean correlation). This is because in the zeroth approximation ρ_* is given by ρ' (Kakushadze, 2014a). *E.g.*, if we have only two alphas α_1 and α_2 and $\Psi_{12} < 0$, we can take $\alpha'_1 \equiv \alpha_1$ and $\alpha'_2 \equiv -\alpha_2$, so the resulting correlation $\Psi'_{12} > 0$.

¹³ Another way of thinking about this is that the first principal component $V^{(1)}$ is closest to the full $SO(N)$ rotational invariance, *i.e.*, to the normalized unit vector $\bar{V}_i \equiv 1/\sqrt{N}$, $i = 1, \dots, N$.

5 Factor Model and Turnover

In this section our goal is to study turnover reduction in the large N limit using the factor model, which provides a simple computational framework for understanding the behavior of the turnover reduction coefficient ρ_* with increasing N . The spectral model formula (10), obtained using the principal component analysis (Kakushadze, 2014a), provides a well-defined prescription for estimating the turnover reduction coefficient ρ_* for a given alpha correlation matrix.¹⁴ However, our goal here is to gain *intuitive* insight into the large N behavior of ρ_* . The factor model approach provides a convenient calculational playground in this regard: we have F risk factors (specified via the $N \times F$ factor loadings matrix Ω_{iA}), the $F \times F$ factor covariance matrix Φ_{AB} , and the diagonal $N \times N$ specific risk matrix Ξ_{ij} . For instance, we can ask: what is the dependence of ρ_* on F in the large N limit?

We will start with a simplified factor model¹⁵ where i) specific risks ξ_i are set to zero, ii) there are no style risk factors, and all cluster risk factors are “binary” in the sense that each α_i belongs to one and only one cluster, and iii) the factor covariance matrix Φ_{AB} is diagonal:¹⁶

$$\xi_i \equiv 0 \quad (17)$$

$$\Omega_{iA} = \delta_{G(i),A} \quad (18)$$

$$\Phi_{AB} = \phi_A \delta_{AB} \quad (19)$$

where

$$G : \{1, \dots, N\} \mapsto \{1, \dots, F\} \quad (20)$$

is the map between alphas and clusters.

We have:

$$\Gamma_{ij} = \phi_{G(i)} \delta_{G(i),G(j)} \quad (21)$$

$$\sigma_i^2 \equiv \Gamma_{ii} = \phi_{G(i)} \quad (22)$$

$$\Psi_{ij} \equiv \frac{1}{\sigma_i \sigma_j} \Gamma_{ij} = \delta_{G(i),G(j)} \quad (23)$$

¹⁴ Even when $M < N$, *i.e.*, when the sample correlation matrix is singular, (10) is still applicable. This is because it uses the first principal component $V^{(1)}$ and the largest eigenvalue $\psi^{(1)}$, whose dependence on M does not alter the results dramatically. One way to see this is to deform a singular correlation matrix using the method discussed in Subsection 3.1 of (Kakushadze, 2014a) based on (Rebonato and Jäckel, 1999) with the non-positive eigenvalues replaced by the smallest positive eigenvalue, as in Section 7. Thus, for the hedge fund data we use in Section 7, for the largest eigenvalue the difference between the undeformed and deformed cases is only about 24%.

¹⁵ We add complexity below, including *non-binary* factor loadings – the simplified factor model serves the purpose of illustrating the key issues without overcomplicating them with unnecessary math. Here we choose binary factor loadings for calculational convenience. As we will see below, the simplified factor model captures the key features of the large N behavior. Also, note that the simplified model of (Kakushadze and Liew, 2014) is a special case of a general one-factor model.

¹⁶ Here the values of the factor loadings elements need not be “binary”. If $\Omega_{iA} = \omega_{iA} \delta_{G(i),A}$ with non-binary ω_{iA} , the result is unchanged, both here and in Subsection 5.2 (non-diagonal factor covariance matrix). What is important here is the binary membership of alphas in clusters.

and the correlation matrix is block-diagonal with F blocks corresponding to F clusters, and each diagonal block has all elements equal 1.

The correlation matrix Ψ_{ij} has $(N - F)$ null eigenvalues¹⁷ and F eigenvalues equal to:

$$\psi^{(A)} = N_A \quad (24)$$

where

$$N_A \equiv \sum_{i=1}^N \delta_{G(i),A} \quad (25)$$

is the number of alphas that belong to the cluster labeled by A . Note that

$$\sum_{A=1}^F N_A = N \quad (26)$$

Also, the right eigenvectors $V_i^{(A)}$ corresponding to the eigenvalues $\psi^{(A)}$ are given by

$$V_i^{(A)} = \frac{1}{\sqrt{N_A}} \delta_{G(i),A} \quad (27)$$

and are normalized such that

$$\sum_{i=1}^N \left(V_i^{(A)} \right)^2 = 1 \quad (28)$$

Note that

$$\rho_* = \left(\frac{N_*}{N} \right)^{\frac{3}{2}} \gtrsim \frac{1}{F^{\frac{3}{2}}} \quad (29)$$

where

$$N_* \equiv \max(N_A, A = 1, \dots, F) \quad (30)$$

Therefore, for fixed F simply increasing N does not reduce turnover indefinitely.

This is because adding more and more alphas that belong to the same clusters does not diversify them as alphas in the same clusters are correlated with each other. To reduce ρ_* one also needs to increase F , *i.e.*, to add alphas that belong to new clusters. Also, note that if F is fixed, adding more and more alphas can decrease ρ_* (unless all alphas are added to the cluster with the largest N_A), but it does not decrease it to zero.

In this regard, let us consider the case where $N \rightarrow \infty$ with $F = \text{fixed}$. From (29) it follows that ρ_* is minimized when N_* is minimized, where N_* is the largest N_A . It then follows that $N_* = \text{ceiling}(N/F)$, where $\text{ceiling}(x) \equiv \lceil x \rceil$ refers to the smallest integer not less than x , and similarly $\text{floor}(x) \equiv \lfloor x \rfloor$ refers to the largest

¹⁷ There is no reason for alarm – these eigenvalues are null because we set the specific risks to zero – see below.

integer not greater than x . Consequently, in the $N \rightarrow \infty$ limit the minimum of ρ_* is at $(\rho_*)_{\min} \equiv F^{-3/2}$. For any finite N the corresponding distribution has F_- values of $N_A = \text{floor}(N/F)$ and F_+ values of $N_A = \text{ceiling}(N/F)$, where $F_- \equiv F - F_+$ and $F_+ \equiv N - F \text{ floor}(N/F)$. Note that $0 \leq F_+ < F$, and in the case where $F_+ = 0$ we have $\text{ceiling}(N/F) = \text{floor}(N/F) = N/F$.

5.1 Effect of Specific Risk

Above we assumed zero specific risk. In this subsection we study effects of non-zero specific risk. To keep things simple, we will assume that

$$\xi_i = \tilde{\xi}_{G(i)} \quad (31)$$

$$\Omega_{iA} = \delta_{G(i),A} \quad (32)$$

$$\Phi_{AB} = \phi_A \delta_{AB} \quad (33)$$

where $\tilde{\xi}_A$, $A = 1, \dots, F$ are some specific risks corresponding to each cluster. *I.e.*, we assume that specific risk is uniform within each cluster, which corresponds to an approximation where specific risk within each cluster is replaced (be it directly or logarithmically), *e.g.*, by mean (or median) specific risk or square root of mean (median) specific variance for that cluster. We do this here for computational simplicity – as we will see, the effect of specific risk on turnover reduction is subleading in the large N limit, and so is the effect of nonuniform specific risk within each cluster.

With the above assumptions we now have:

$$\Gamma_{ij} = \tilde{\xi}_{G(i)}^2 \delta_{ij} + \phi_{G(i)} \delta_{G(i),G(j)} \quad (34)$$

$$\sigma_i^2 \equiv \Gamma_{ii} = \tilde{\xi}_{G(i)}^2 + \phi_{G(i)} \quad (35)$$

$$\Psi_{ij} \equiv \frac{1}{\sigma_i \sigma_j} \Gamma_{ij} = \frac{\tilde{\xi}_{G(i)}^2}{\tilde{\xi}_{G(i)}^2 + \phi_{G(i)}} \delta_{ij} + \frac{\phi_{G(i)}}{\tilde{\xi}_{G(i)}^2 + \phi_{G(i)}} \delta_{G(i),G(j)} \quad (36)$$

This correlation matrix has the following eigenvalue structure. For each $A = 1, \dots, F$, it has $(N_A - 1)$ eigenvalues equal

$$\tilde{\psi}^{(A)} = \frac{\tilde{\xi}_A^2}{\tilde{\xi}_A^2 + \phi_A}, \quad \tilde{d}_A = N_A - 1, \quad A = 1, \dots, F \quad (37)$$

where \tilde{d}_A is the degeneracy of each such eigenvalue. This gives total of $N - F$ eigenvalues. The remaining F eigenvalues (each with unit degeneracy $d_A = 1$) are given by:

$$\psi^{(A)} = \frac{\tilde{\xi}_A^2 + N_A \phi_A}{\tilde{\xi}_A^2 + \phi_A}, \quad d_A = 1, \quad A = 1, \dots, F \quad (38)$$

This reduces to the previous result when all $\tilde{\xi}_A \equiv 0$.

The right eigenvectors $V_i^{(A)}$ corresponding to the eigenvalues $\psi^{(A)}$, as before, are given by

$$V_i^{(A)} = \frac{1}{\sqrt{N_A}} \delta_{G(i),A} \quad (39)$$

So we have:

$$\rho_* = \frac{1 + \frac{1}{N_*} \zeta_*}{1 + \zeta_*} \left(\frac{N_*}{N} \right)^{\frac{3}{2}} \quad (40)$$

where

$$\zeta_A \equiv \frac{\tilde{\xi}_A^2}{\phi_A} \quad (41)$$

and $\zeta_* \equiv \zeta_A$ for the value of A for which $N_A = N_*$ (assuming for the sake of simplicity that N_A are all unique). As we see, specific risk does not affect the large N behavior of ρ_* ; instead, it simply amounts to reducing the overall coefficient in ρ_* , but does not affect the conclusions we arrived at in the zero specific risk case.

In this regard, *assuming* it is kept finite (see below), specific risk does not qualitatively affect turnover reduction in the large N limit, albeit it is important, *e.g.*, in the weight optimization even in the large N limit. Furthermore, above we assumed that specific risk is uniform across each cluster, but relaxing this assumption does not change the above conclusions relating to turnover reduction in the large N limit.

Here the following clarifying remark is in order. As mentioned above, specific risk does further reduce the overall turnover reduction coefficient. If we take $\zeta_* \rightarrow \infty$, *i.e.*, the factor risk is negligible compared with specific risk, then turnover goes to zero. Indeed, in this case we simply have N uncorrelated alphas and in the large N limit we have $\rho_* \sim \sqrt{N_*/N^3} \rightarrow 0$. In the above analysis, when discussing the behavior of ρ_* in the large N limit, we assume that $N \rightarrow \infty$ (and, consequently, $N_* \rightarrow \infty$) with $\zeta_* = \text{fixed}$. Then we have $\rho_* \sim (N_*/N)^{3/2}/(1 + \zeta_*) \rightarrow \text{const}$.

5.2 Non-diagonal Factor Covariance Matrix

Another simplifying assumption we made above was that the factor covariance matrix is diagonal. In this subsection we relax this assumption. Since we already know that specific risk does not affect the qualitative picture, we will set it to zero in order not to overcomplicate things, but we will make no assumption about the factor covariance matrix:

$$\xi_i \equiv 0 \quad (42)$$

$$\Omega_{iA} = \delta_{G(i),A} \quad (43)$$

We now have:

$$\Gamma_{ij} = \Phi_{G(i),G(j)} \quad (44)$$

$$\sigma_i^2 \equiv \Gamma_{ii} = \Phi_{G(i),G(i)} \quad (45)$$

$$\Psi_{ij} \equiv \frac{1}{\sigma_i \sigma_j} \Gamma_{ij} = \hat{\Psi}_{G(i),G(j)} \quad (46)$$

where $\widehat{\Psi}_{AB}$ is the factor correlation matrix:

$$\Phi_{AB} \equiv \sqrt{\Phi_{AA}} \sqrt{\Phi_{BB}} \widehat{\Psi}_{AB} \quad (47)$$

Then the generalization of the eigenvectors (27) for non-diagonal Φ_{AB} can be found as follows.

There are F such eigenvectors. They are of the form

$$V_i^{(A)} = \frac{1}{\sqrt{N_{G(i)}}} \chi_{G(i)}^{(A)} \quad (48)$$

Note that

$$\sum_{j=1}^N \Psi_{ij} V_j^{(A)} = \sum_{j=1}^N \widehat{\Psi}_{G(i),G(j)} \frac{1}{\sqrt{N_{G(j)}}} \chi_{G(j)}^{(A)} = \frac{1}{\sqrt{N_{G(i)}}} \sum_{B=1}^F \widehat{\Psi}'_{G(i),B} \chi_B^{(A)} \quad (49)$$

where

$$\widehat{\Psi}'_{AB} \equiv \sqrt{N_A} \widehat{\Psi}_{AB} \sqrt{N_B} \quad (50)$$

Note that $\widehat{\Psi}'$ is a symmetric matrix and $\chi_B^{(A)}$ are the eigenvectors of $\widehat{\Psi}'$:

$$\sum_{B=1}^F \widehat{\Psi}'_{CB} \chi_B^{(A)} = \widehat{\psi}^{(A)} \chi_C^{(A)} \quad (51)$$

where $\widehat{\psi}^{(A)}$ are the eigenvalues of $\widehat{\Psi}'$.

So we have

$$\sum_{j=1}^N \Psi_{ij} V_j^{(A)} = \widehat{\psi}^{(A)} V_i^{(A)} \quad (52)$$

and

$$1 = \sum_{i=1}^N \left(V_i^{(A)} \right)^2 = \sum_{B=1}^F \left(\chi_B^{(A)} \right)^2 \quad (53)$$

fixes the normalization of $\chi_B^{(A)}$.

Next, note that the diagonal elements $\widehat{\Psi}_{AA} \equiv 1$, and (in matrix notation)

$$\widehat{\Psi}' = Q \widehat{\Psi} Q \quad (54)$$

where

$$Q \equiv \text{diag} \left(\sqrt{N_A} \right) \quad (55)$$

is a diagonal $F \times F$ matrix. This implies that, since $\widehat{\Psi}_{AB}$ is positive definite, so is $\widehat{\Psi}'_{AB}$. Furthermore, we have:

$$\text{Tr} \left(\widehat{\Psi}' \right) = \sum_{A=1}^F N_A = N \quad (56)$$

This implies that the off-diagonal elements of $\widehat{\Psi}_{AB}$ have the following effect on the eigenvalues $\widehat{\psi}^{(A)}$. When the off-diagonal elements of $\widehat{\Psi}_{AB}$ are zero, these eigenvalues are equal N_A . When they are nonzero, these eigenvalues generally are different from N_A , but they are still positive and their sum is still equal N :

$$\forall A = 1, \dots, F : \quad \widehat{\psi}^{(A)} > 0 \quad (57)$$

$$\sum_{A=1}^F \widehat{\psi}^{(A)} = N \quad (58)$$

Also

$$\left| \sum_{i=1}^N V_i^{(A)} \right| = \left| \sum_{B=1}^F \sqrt{N_B} \chi_B^{(A)} \right| \quad (59)$$

This implies that (V_i^* is the eigenvector corresponding to $\widehat{\psi}^*$)

$$\widehat{\psi}^* \equiv \max \left(\widehat{\psi}^{(A)}, A = 1, \dots, F \right) \gtrsim \frac{N}{F} \quad (60)$$

$$\left| \sum_{i=1}^N V_i^* \right| \gtrsim \sqrt{\frac{N}{F}} \quad (61)$$

and the above conclusions relating to turnover reduction in the large N limit are unchanged even if the factor covariance matrix Φ_{AB} is not diagonal – the off-diagonal elements generally increase the lower bound. Appendix A discusses (61) in more detail. Adding specific risk does not modify this result.

5.3 Effect of Style Factors

Adding style risk factors into the mix does not alter the above conclusions. *Assuming* there is a finite and small number of style risk factors,¹⁸ their inclusion has subleading effect when the number of cluster risk factors is large.

A quick way to see this is to consider a simple factor model with zero specific risk and a single style factor Ω_i . The 1×1 factor covariance matrix can be absorbed into the definition of Ω_i . We then have

$$\Gamma_{ij} = \Omega_i \Omega_j \quad (62)$$

$$\sigma_i^2 \equiv \Gamma_{ii} = \Omega_i^2 \quad (63)$$

$$\Psi_{ij} \equiv \frac{1}{\sigma_i \sigma_j} \Gamma_{ij} = 1 \quad (64)$$

¹⁸ This is not an unreasonable assumption. In the case of stocks, the number of style risk factors is substantially smaller than the number of industry risk factors. In the case of alphas, it appears that the number of basic style risk factors should be even smaller, albeit there is a trick to effectively increase their number – see (Kakushadze, 2014b) for details.

This matrix has $(N - 1)$ null eigenvalues and a single nonzero eigenvalue equal N . Adding a few more style factors can reduce the maximum eigenvalue of the correlation matrix only by a factor of order 1. It takes a large number of risk factors, be it binary or otherwise, to reduce the maximum eigenvalue of the correlation matrix substantially, which is what we found above.

5.4 Non-binary Factor Loadings

Above, for the sake of simplicity, we assumed that the factor loadings Ω_{iA} are binary. In this subsection we consider the case of arbitrary non-binary factor loadings. In this case it is convenient to absorb the factor covariance matrix into the definition of the factor loadings:

$$\Gamma = \Xi + \tilde{\Omega} \tilde{\Omega}^T \quad (65)$$

$$\tilde{\Omega} \equiv \Omega \tilde{\Phi} \quad (66)$$

$$\tilde{\Phi} \tilde{\Phi}^T = \Phi \quad (67)$$

where $\tilde{\Phi}_{AB}$ is the Cholesky decomposition of Φ_{AB} , which is assumed to be positive-definite.

For the sake of simplicity, as we did before Subsection 5.1, let us set specific risk to zero – we comment on the effect of specific risk below. Then the covariance matrix reads:

$$\Gamma_{ij} = \sum_{A=1}^F \tilde{\Omega}_{iA} \tilde{\Omega}_{jA} \quad (68)$$

$$\sigma_i^2 \equiv \Gamma_{ii} = \sum_{A=1}^F \tilde{\Omega}_{iA}^2 \quad (69)$$

$$\Psi_{ij} \equiv \frac{1}{\sigma_i \sigma_j} \Gamma_{ij} = \sum_{A=1}^F \Lambda_{iA} \Lambda_{jA} \quad (70)$$

where

$$\Lambda_{iA} \equiv \frac{1}{\sigma_i} \tilde{\Omega}_{iA} \quad (71)$$

The correlation matrix Ψ_{ij} has $N - F$ null eigenvalues and F non-zero eigenvalues.¹⁹ The latter are given by the eigenvalues $\psi^{(A)}$ of the matrix

$$Q_{AB} \equiv \sum_{i=1}^N \Lambda_{iA} \Lambda_{iB} \quad (72)$$

whose properties also determine the corresponding eigenvectors of Ψ_{ij} .

¹⁹ More precisely, these F eigenvalues are nonzero assuming that the matrix Q_{AB} defined below is nonsingular – see below.

Let

$$Q = W Z W^T \quad (73)$$

$$Z = \text{diag}(\psi^{(A)}) \quad (74)$$

where W is the $F \times F$ matrix, normalized such that $W^T W = 1$, whose columns are the right eigenvectors of Q_{AB} . We can assume that all $\psi^{(A)} > 0$, $A = 1, \dots, F$ – the matrix Q_{AB} is positive-definite provided that the N -vectors $y_i^{(A)} \equiv \Lambda_{iA}$, $A = 1, \dots, F$ are linearly independent, which we assume to be the case without loss of generality. Further, let

$$V_i^{(A)} \equiv \frac{1}{\sqrt{\psi^{(A)}}} \sum_{B=1}^F \Lambda_{iB} W_{BA} \quad (75)$$

Then we have

$$\sum_{j=1}^N \Psi_{ij} V_j^{(A)} = \psi^{(A)} V_i^{(A)} \quad (76)$$

$$\sum_{i=1}^N V_i^{(A)} V_i^{(B)} = \delta_{AB} \quad (77)$$

So, $V_i^{(A)}$ are the properly normalized F eigenvectors of Ψ_{ij} corresponding to the eigenvalues $\psi^{(A)}$, which have the following properties:

$$\sum_{A=1}^N \psi^{(A)} = \text{Tr}(Q) = \text{Tr}(\Psi) = N \quad (78)$$

$$\psi^* \equiv \max(\hat{\psi}^{(A)}, A = 1, \dots, F) \gtrsim \frac{N}{F} \quad (79)$$

We still need to estimate

$$\left| \sum_{i=1}^N V_i^* \right| \quad (80)$$

where V_i^* is the eigenvector corresponding to the largest eigenvalue ψ^* . Details are relegated to Appendix B – thus, in the case of non-binary factor loadings we also find a finite limit for the turnover reduction coefficient in the large N limit. Adding specific risk does not modify this result.

6 Practical Application

As we saw above, in the context of a factor model for alphas, turnover reduction does not necessarily go to zero by adding more and more alphas. One needs to add more and more *different* types of new alphas to achieve indefinite turnover

reduction. In the factor model context this translates into adding alphas that form new clusters, not adding alphas into old clusters. Intuitively, this result is not surprising – to achieve turnover reduction, one needs to add more and more alphas that are (almost) uncorrelated with the existing alphas. However, the factor model framework provides a way of quantifying this statement. In particular, it allows to design a simple test of whether new alphas have a potential for turnover reduction.

Thus, suppose we have N alphas α_i that form F clusters with the factor loadings Ω_{iA} , $i = 1, \dots, N$, $A = 1, \dots, F$. Suppose new alphas are developed and we need to assess whether they have a potential for turnover reduction. The question is whether these new alphas form a new cluster.²⁰ For simplicity let us consider the case where the claim is that the new alphas are all such that they form a single new cluster. Let the old plus new alphas be $\alpha'_{i'}$, $i' = 1, \dots, N'$, where N' is the total number of the old plus new alphas and $(N' - N)$ is the number of the new alphas. Let the new factor loadings matrix be $\Omega'_{i'A'}$, $A' = 1, \dots, F'$, where $F' = F + 1$.

Now we can run two regressions (without intercept), first α_i over Ω_{iA} , and second $\alpha'_{i'}$ over $\Omega'_{i'A'}$. In R notations:

$$\alpha \sim -1 + \Omega \tag{81}$$

$$\alpha' \sim -1 + \Omega' \tag{82}$$

One can now compare the F-statistic for each of these regressions. In actuality, α_i and $\alpha'_{i'}$ are time series: $\alpha_i(t_s)$ and $\alpha'_{i'}(t_s)$, $s = 0, 1, \dots, M$. So one can look, *e.g.*, at two time-series vectors of F-statistic and assess whether the hypothesis that new alphas form a new cluster has better overall F -statistic.²¹

The above discussion assumes that the factor loadings Ω_{iA} corresponding to the existing clusters are known. The question is, how does one identify these existing clusters in the first place? A somewhat “primitive”, albeit perhaps practical, way of building out a binary alpha classification is to use the method of this section recursively. If there are many independent alpha sources (developers), in the zeroth approximation they effectively label the clusters. Every time new alphas are developed, one needs to filter out those too highly correlated with the existing alphas. If one has a binary classification with K clusters, then one can build a classification with $K + 1$ clusters, and repeat this recursively. Most care is needed when K is low – once K is large, there is more statistics available for filtering out redundant new alphas. To improve this method at low K , one can supplement it with algorithms mentioned in footnote 10, where binary factor loadings are obtained using sample correlation matrix and/or principal components.²²

²⁰ As mentioned in Subsection 5.1, specific risk decreases turnover. If new alphas belong to an existing cluster but have higher specific risk, turnover could be reduced. However, specific risk is computed in the risk factor model *after* identifying all relevant risk factors, including any clusters.

²¹ In comparing the two F-statistic vectors, one can remove (or smooth via standard techniques, *e.g.*, Winsorization) outliers to improve statistical significance of the comparison.

²² Such (again, typically proprietary) algorithms are usually based on ranking techniques. They are expected to work better precisely when K is low, as there is a practical bound on M .

7 “Power Law” and Number of Clusters

Our main result above is that, using a factor model approach, we arrived at the “power law” (29), which suggests that the turnover reduction coefficient ρ_* is controlled the number of clusters F :

$$\rho_* \gtrsim \frac{1}{F^{\frac{3}{2}}} \quad (83)$$

Here we ask two questions. First, can we estimate F for a given correlation matrix Ψ_{ij} ? Second, can we increase F indefinitely by increasing N , or is there a limit?

To answer the first question, let us first recall from Eq. (60) that

$$\psi^* \gtrsim \frac{N}{F} \quad (84)$$

This gives us a simple method for estimating F :

$$F \gtrsim \frac{N}{\psi^*} \quad (85)$$

Here the following remark is in order. The lower bound (85) is applicable assuming $N \lesssim M$, so that the correlation matrix is nonsingular. If $M < N$, then the correlation matrix is singular and the number of clusters $F \leq M$, so depending on M the lower bound (85) may or may not be informative. As an *illustrative*²³ example, let us use the same Morningstar data as in Fig.1 of (Kakushadze and Liew, 2014) and Fig.1 of (Kakushadze, 2014a), which is the data for 1990-2014 for $N = 657$ monthly hedge fund returns (HF). For the undeformed²⁴ correlation matrix of the raw HF we have $\psi^* \approx 207$ and $F \gtrsim 3.17$, and for the corresponding deformed correlation matrix of the raw HF we have $\psi^* \approx 158$ and $F \gtrsim 4.15$. For the undeformed correlation matrix of the residuals (plus the intercepts, which have no effect) of the HF adjusted for RF (whose effect is small) and regressed over Mkt-RF and the Fama-French risk factors SMB, HML, WML, we have $\psi^* \approx 93.9$ and $F \gtrsim 7.00$, and for the corresponding deformed correlation matrix we have $\psi^* \approx 71.1$ and $F \gtrsim 9.24$. Note how regressing away Mkt-RF and the Fama-French risk factors, which evidently dominate in the raw HF, improves the lower bound. Also note that in this example M is substantially larger than the lower bound we obtained, so the latter is informative.

²³ We emphasize the adjective “illustrative” for the reason that, because various hedge funds in this data do/did not all trade the same underlying instruments and also the corresponding time series are not 100% overlapping (some hedge funds are dead, some are newer than others, *etc.*), it would not necessarily be correct to assume that their trades could be crossed. Therefore, we use this data only to *illustrate* various properties of the correlation matrix, and not necessarily to directly draw any conclusions about turnover reduction had these alpha streams actually crossed their trades.

²⁴ The raw HF have non-uniform N /As, so the correlation matrix is computed by omitting such pair-wise N /As and, as mentioned in footnote 6, has negative eigenvalues, which are dealt with by deforming the correlation matrix using the method discussed in Subsection 3.1 of (Kakushadze, 2014a) based on (Rebonato and Jäckel, 1999) with the non-positive eigenvalues replaced by the smallest positive eigenvalue (which is not a zero distorted by computational rounding) – also see the end of Section 7.1. Here the values we give for ψ^* and F are rounded to 2 significant figures.

7.1 Regression Correlations

Above we discussed how to obtain a lower bound for the number of clusters. A nontrivial method for estimating an upper bound for F is as follows. Let Λ_{iA} , $A = 1, \dots, K$ be the $N \times K$ matrix of the first K principal components of Ψ_{ij} , *i.e.*, the columns of Λ_{iA} are the first K principal components.²⁵ Next, let $\epsilon_i(t_s)$ be the time series of the regression residuals, which in R notations reads (*i.e.*, the regression is without intercept and with trivial weights):

$$\tilde{\alpha} \sim -1 + \Lambda \quad (86)$$

where (see Section 2)

$$\tilde{\alpha}_i \equiv \frac{\alpha_i}{\sigma_i} \quad (87)$$

Note that

$$\text{Cov}(\tilde{\alpha}_i, \tilde{\alpha}_j) = \text{Cor}(\alpha_i, \alpha_j) = \Psi_{ij} \quad (88)$$

In matrix notations we have:

$$\epsilon(t_s) = (1 - Y) \tilde{\alpha}(t_s) \quad (89)$$

where

$$Y \equiv \Lambda (\Lambda^T \Lambda)^{-1} \Lambda^T \quad (90)$$

is a projection matrix: $Y^2 = Y$.

Note that:

$$\text{Cov}(\epsilon_i, \epsilon_j) = [(1 - Y) \Psi (1 - Y)]_{ij} \equiv \Phi_{ij} \quad (91)$$

Let

$$\xi_i^2 \equiv \Phi_{ii} \quad (92)$$

$$\tilde{\epsilon}_i \equiv \frac{\epsilon_i}{\xi_i} \quad (93)$$

Then

$$\tilde{\Phi}_{ij} \equiv \text{Cov}(\tilde{\epsilon}_i, \tilde{\epsilon}_j) = \text{Cor}(\epsilon_i, \epsilon_j) = \frac{1}{\xi_i \xi_j} \Phi_{ij} \quad (94)$$

I.e., $\tilde{\Phi}_{ij}$ is the correlation matrix of the regression residuals ϵ_i . Moreover, to compute it, we do not need to know α_i ; we only need to know the correlation matrix Ψ_{ij} .

Let

$$\zeta_1 \equiv \text{Mean}(\tilde{\Phi}_{ij}, i \neq j) \quad (95)$$

$$\zeta_2 \equiv \text{Median}(\tilde{\Phi}_{ij}, i \neq j) \quad (96)$$

²⁵ There are other ways of constructing Λ_{iA} , but for our purposes here building it out of the principal components suffices.

By looking at ζ_1 and ζ_2 as a function of K , we can determine the value of K above which ζ_1 and ζ_2 no longer decrease substantially. This value of K estimates (the upper bound for) the number of clusters F for the correlation matrix Ψ_{ij} . For the aforementioned case of the HF data the graphs for ζ_1 and ζ_2 as functions of K are given by Fig.1 and Fig.2. From these graphs we can deduce that F is between 5 and 10, which is consistent with the above lower bound results based on the largest eigenvalue method. Note that the latter method is based on a binary cluster assumption, whereas the former method is based on non-binary analysis using principal components. These complementary methods can be used to estimate the number of clusters in the case of much larger N . Also, note that to use the principal component method, the correlation matrix must be positive-definite, so in Fig.1 and Fig.2 we deformed the correlation matrix using the method discussed in Subsection 3.1 of (Kakushadze, 2014a) based on (Rebonato and Jäckel, 1999) with the non-positive eigenvalues replaced by the smallest positive eigenvalue (see footnote 24).

8 Is There a Limit to Turnover Reduction?

The answer is yes – assuming the number of underlying tradable instruments is finite. Indeed, in this case the investment level is finite, and if we assume that by adding more and more alphas, *i.e.*, by taking $N \rightarrow \infty$, we can reduce the turnover indefinitely, we would arrive at a *static* portfolio of finite underlying tradable instruments. A dollar-neutral (or, more generally, “market-neutral”) static portfolio cannot have high Sharpe and return. Long-only static portfolios with positive returns do exist – S&P is an example, albeit its Sharpe ratio is low.²⁶ However, a dollar-neutral static portfolio with high Sharpe and return cannot persist for too long. Suppose it does. Let it have I long and I short dollar positions. Then by adding I dollar long S&P, we can synthesize a net I dollar long portfolio with the following P&L, volatility and Sharpe:

$$P = P_1 + P_2 \tag{97}$$

$$R = \sqrt{R_1^2 + R_2^2 + 2 \rho R_1 R_2} \tag{98}$$

$$S = \frac{P}{R} \tag{99}$$

where P_1 and P_2 are the P&Ls of the dollar-neutral portfolio and S&P, R_1 and R_2 are the corresponding volatilities, and ρ is the correlation between the P&Ls, which can be assumed to be low: $|\rho| \ll 1$. Then it is evident that, if the dollar-neutral portfolio significantly outperforms S&P, the synthetic long portfolio, which is static, outperforms S&P. However, such a condition cannot last long as it will be quickly arbitrated away – the synthetic portfolio is static, hence no apparent obstruction.

²⁶ More precisely, even S&P is not static, it is quasi-static due to periodic re-balancing, but such effects are not important for our discussion here. Also, it is not important here whether a return is positive with respect to a non-negative benchmark.

The above discussion implies that there is a limit to how much turnover can be reduced. This, in turn, implies that the number of clusters is finite so long as the number of underlying tradable instruments is finite. Once this limit is reached – and the number of clusters for any given set of N alphas can be estimated using the methods discussed in Section 7 – the development of more and more alphas no longer serves the purpose of reducing the turnover, but rather of improving the return of the portfolio by i) recycling away the alphas with degraded performance and ii) effectively enlarging the weight space and thereby improving the alpha weight optimization. Since the number of clusters is finite, they should be determined and used in order to optimize the search for new alphas to maximize the impact of i) and ii) above. And the large N limit – as in theoretical physics ('t Hooft, 1974) – proves to be a powerful tool for understanding turnover reduction and other aspects of portfolios consisting of a large number N of alphas.

9 Comments

We end this note with a few clarifying comments. First, (Kakushadze and Liew, 2014), (Kakushadze, 2014a) and this paper assume that all alphas are traded on the same execution platform. In practice this means that the N alphas α_i are combined with some weights w_i (typically, via optimization (Kakushadze, 2014c)) into a single alpha, and it is this combined alpha that is traded. This is the efficient way; the internal crossing is automatic. Second, as discussed in more detail in (Kakushadze and Liew, 2014) and (Kakushadze, 2014a), generally, it is no easy feat to precisely describe internal crossing and turnover reduction. In a portfolio consisting of a large number of underlying tradable instruments (*e.g.*, stocks), precise details of internal crossing depend on the detailed portfolio position and trade data. Modeling turnover reduction via alpha correlations is exactly that – *modeling*. *E.g.*, correlations between trades and correlations between positions are not the same. However, as argued in (Kakushadze, 2014a), notwithstanding the caveats, the spectral model is expected to be a good approximation in the large N limit for reasonably distributed individual turnovers and weights. In this regard, alphas being optimized, *i.e.*, combined into a single alpha, makes for a significant simplification – this avoids the issue of timing trades between different alphas, which is what would have to be done had the alphas actually been traded individually. Furthermore, in the large N limit the leading contribution (in the $1/N$ expansion) into the turnover reduction coefficient has a correlation-like structure (Kakushadze and Liew, 2014), and, as was argued in (Kakushadze, 2014a), this leading contribution is well-approximated (under the aforementioned conditions) by the contribution from the first principal component of the alpha correlation matrix. Third, as mentioned above, the use of the factor model approach and (binary) clusters is a convenient computational tool, which allows us to gain *intuitive* insight into turnover reduction beyond and complementary to the principal component approach of (Kakushadze, 2014a).

A Properties of Eigenvectors

Here we elaborate on (61) we used in Subsection 5.2, which may not be evident. To do this, instead of dealing with general factor correlation matrix $\widehat{\Psi}_{AB}$, here we will study a simpler case where

$$\widehat{\Psi}_{AB} \equiv \rho, \quad A \neq B \quad (100)$$

Let

$$\widetilde{\chi}^{(A)} \equiv \sum_{B=1}^F \sqrt{N_B} \chi_B^{(A)} \quad (101)$$

Then from (51) we have:

$$\chi_C^{(A)} = \rho \widetilde{\chi}^{(A)} \frac{\sqrt{N_C}}{\widehat{\psi}^{(A)} - (1 - \rho) N_C} \quad (102)$$

Multiplying both sides by $\sqrt{N_C}$ and summing over $C = 1, \dots, F$, we obtain:

$$\rho \sum_{C=1}^F \frac{N_C}{\widehat{\psi}^{(A)} - (1 - \rho) N_C} = 1 \quad (103)$$

The eigenvalues $\widehat{\psi}^{(A)}$ are the F roots of this equation (assuming for the sake of simplicity that all N_C are distinct). It further follows from (103) that

$$\rho \sum_{C=1}^F \frac{1}{\widehat{\psi}^{(A)} - (1 - \rho) N_C} = \frac{\phi^*}{\widehat{\psi}^{(A)}} \quad (104)$$

where

$$\phi^* \equiv 1 + \rho (F - 1) \quad (105)$$

is the non-degenerate eigenvalue of $\widehat{\Psi}_{AB}$ – the other $(F - 1)$ eigenvalues $\phi' \equiv 1 - \rho$.

For $A \neq B$ we have

$$\begin{aligned} \sum_{C=1}^F \chi_C^{(A)} \chi_C^{(B)} &= \rho^2 \widetilde{\chi}^{(A)} \widetilde{\chi}^{(B)} \sum_{C=1}^F \frac{N_C}{\left[\widehat{\psi}^{(A)} - (1 - \rho) N_C \right] \left[\widehat{\psi}^{(B)} - (1 - \rho) N_C \right]} = \\ &= \frac{\rho^2 \widetilde{\chi}^{(A)} \widetilde{\chi}^{(B)}}{(1 - \rho) \left(\widehat{\psi}^{(B)} - \widehat{\psi}^{(A)} \right)} \sum_{C=1}^F \left[\frac{\widehat{\psi}^{(A)}}{\widehat{\psi}^{(A)} - (1 - \rho) N_C} - \frac{\widehat{\psi}^{(B)}}{\widehat{\psi}^{(B)} - (1 - \rho) N_C} \right] = \\ &= 0 \end{aligned} \quad (106)$$

So different eigenvectors are orthogonal to each other, as they should be.

From (53) we have:

$$1 = \sum_{C=1}^F \left(\chi_C^{(A)} \right)^2 = \rho^2 \left(\widetilde{\chi}^{(A)} \right)^2 \sum_{C=1}^F \frac{N_C}{\left[\widehat{\psi}^{(A)} - (1 - \rho) N_C \right]^2} \quad (107)$$

The last sum can be tackled as follows. Let ∂ denote the derivative w.r.t. ρ , *e.g.*:

$$\partial \widehat{\psi}^{(A)} \equiv \frac{\partial \widehat{\psi}^{(A)}}{\partial \rho} \quad (108)$$

Differentiating (104) w.r.t. ρ and rearranging terms, we obtain:

$$\rho^2 \left[\widehat{\psi}^{(A)} + (1 - \rho) \partial \widehat{\psi}^{(A)} \right] \sum_{C=1}^F \frac{N_C}{\left[\widehat{\psi}^{(A)} - (1 - \rho) N_C \right]^2} = \phi^* - \rho \partial \phi^* \quad (109)$$

Together with (105) and (107) this gives the following simple expression:

$$\left(\widetilde{\chi}^{(A)} \right)^2 = \widehat{\psi}^{(A)} + (1 - \rho) \partial \widehat{\psi}^{(A)} \quad (110)$$

For our purposes here we will not need the explicit form of $\partial \widehat{\psi}^{(A)}$.

We are interested in understanding the behavior of

$$\left| \sum_{i=1}^N V_i^* \right| = |\widetilde{\chi}^*| \quad (111)$$

where $\widetilde{\chi}^* \equiv \widetilde{\chi}^{(A)}$ for the value of A for which $\widehat{\psi}^{(A)} = \widehat{\psi}^*$. The largest eigenvalue monotonically increases as ρ increases from 0 to 1, *i.e.*, $\partial \widehat{\psi}^* > 0$. This then implies (61) as we have (60). Also, note that $\det(\widehat{\Psi}') = \det(\widehat{\Psi}) \prod_{A=1}^F N_A$, which implies that as ρ approaches 1, $(F - 1)$ eigenvalues go to zero, while the largest eigenvalue goes to N – recall that $\widehat{\Psi}$ has one eigenvalue equal $\phi^* = 1 + (F - 1) \rho$ and $(F - 1)$ eigenvalues equal $\phi' = 1 - \rho$.

Fig.3 is a graph of $\widehat{\psi}^*$ *vs.* ρ for a randomly constructed matrix with $F = 50$ and $N = 2061$. Also, $F = 2$ provides analytical insight into the eigenvalue structure. Let $N_1 > N_2$. The eigenvalues are given by

$$\widetilde{\psi}^{(1)} = \frac{1}{2} \left[N_1 + N_2 + \sqrt{(N_1 - N_2)^2 + 4 N_1 N_2 \rho^2} \right] \quad (112)$$

$$\widetilde{\psi}^{(2)} = \frac{1}{2} \left[N_1 + N_2 - \sqrt{(N_1 - N_2)^2 + 4 N_1 N_2 \rho^2} \right] \quad (113)$$

For $\rho = 0$ the eigenvalues are N_1 and N_2 . The larger eigenvalue monotonically increases to $N = N_1 + N_2$ as ρ increases from 0 to 1, while the lower eigenvalue decreases to 0.

B Non-binary Case

Here our goal is to estimate (80). To do this, let (here we use the notations of Subsection 5.4)

$$\Lambda_{iA} \equiv \lambda_A + \widetilde{\Lambda}_{iA} \quad (114)$$

$$\sum_{i=1}^N \widetilde{\Lambda}_{iA} \equiv 0 \quad (115)$$

I.e., $\tilde{\Lambda}_{iA}$ are obtained from Λ_{iA} by demeaning its columns. We have

$$\zeta_A \equiv \sum_{i=1}^N V_i^{(A)} = \frac{N}{\sqrt{\psi^{(A)}}} \sum_{B=1}^F \lambda_B W_{BA} \quad (116)$$

$$Q_{AB} = N \lambda_A \lambda_B + \tilde{Q}_{AB} \quad (117)$$

$$\tilde{Q}_{AB} \equiv \sum_{i=1}^N \tilde{\Lambda}_{iA} \tilde{\Lambda}_{iB} \quad (118)$$

Let

$$\tilde{Q} = \tilde{W} \tilde{Z} \tilde{W}^T \quad (119)$$

$$\tilde{Z} = \text{diag}(q_A) \quad (120)$$

where \tilde{W} is the $F \times F$ matrix, normalized such that $\tilde{W}^T \tilde{W} = 1$, whose columns are the right eigenvectors of \tilde{Q}_{AB} corresponding to the eigenvalues q_A . Note that, up to a similarity transformation, Q_{AB} is given by a one-factor model:

$$Q = \tilde{W} \hat{Q} \tilde{W}^T \quad (121)$$

$$\hat{Q} \equiv q_A \delta_{AB} + \hat{\lambda}_A \hat{\lambda}_B \quad (122)$$

$$\hat{\lambda}_A \equiv \sqrt{N} \sum_{B=1}^F \lambda_B \tilde{W}_{BA} \quad (123)$$

Instead of dealing with the most general case, here we will assume that the eigenvalues q_A are identical: $q_A \equiv q$. This can be thought of as an approximation where volatilities q_A in the above one-factor model are replaced (be it directly or logarithmically), *e.g.*, by mean (or median) volatility.²⁷ This approximation will suffice for our purposes here simplifying the math to the point where it is illuminating. We then have

$$Q_{AB} = N \lambda_A \lambda_B + q \delta_{AB} \quad (124)$$

$$\psi^* = N \chi^2 + q \quad (125)$$

$$\chi \equiv \sqrt{\sum_{A=1}^F \lambda_A^2} \quad (126)$$

$$\psi' = q \quad (127)$$

$$W_{A*} = \frac{\lambda_A}{\chi} \quad (128)$$

Here ψ^* is the largest eigenvalue of Q_{AB} , while ψ' are the other (degenerate) $F - 1$ eigenvalues. Also, W_{A*} is the right eigenvector of Q_{AB} corresponding to ψ^* . Thus:

$$\zeta_* = \frac{N}{\sqrt{\psi^*}} \sum_{B=1}^F \lambda_B W_{B*} = \frac{N \chi}{\sqrt{\psi^*}} \quad (129)$$

²⁷ This is analogous to our simplifying approximation in Subsection 5.1, where we assumed uniform specific risk across each cluster.

Here we need to understand the possible values of χ and q . First, note that χ cannot vanish. Indeed, if it does, then all $\lambda_A \equiv 0$, which would then imply that $\sum_{j=1}^N \Psi_{ij} \equiv 0$, $i = 1, \dots, N$. This would imply that at least some off-diagonal elements of Ψ_{ij} are negative. Let $\Psi_{i_1 j_1} < 0$ for some $i_1 \neq j_1$. Then, by flipping the sign of the alpha corresponding to j_1 , $\alpha'_{j_1} \equiv -\alpha_{j_1}$ (and keeping all other alphas unchanged, $\alpha'_j \equiv \alpha_j$, $j \neq j_1$), we would get the corresponding correlation matrix Ψ'_{ij} such that $\sum_{j=1}^N \Psi'_{i_1 j} = -2 \Psi_{i_1 j_1} > 0$. So, $\chi > 0$. Also, recall that $\text{Tr}(Q) = N$, which gives $N \chi^2 + F q = N$, so $q < N/F$ and $\psi^* = N \chi^2 + q = N - (F-1) q > N/F$, which is consistent with (79). However, there is a stronger bound on χ . A simple argument goes as follows. Note that $\sum_{i,j=1}^N \Psi_{ij} = N^2 \sum_{A=1}^N \lambda_A^2 = N^2 \chi^2$. On the other hand, $\sum_{i,j=1}^N \Psi_{ij} = N^2 \rho' = N(1 + (N-1)\bar{\rho})$, where $\bar{\rho}$ is the mean correlation (see footnote 12). On general grounds (Kakushadze, 2014a), we expect that $\rho_* = \gamma \rho'$, where $\gamma \sim 1$ (and typically $\gamma > 1$). It then follows that $\chi \approx \sqrt{\rho_*/\gamma}$ and since $\rho_* = \psi^* \zeta_*/N^{3/2}$, where ζ_* is given by (129), we get

$$\rho_* \approx \frac{\psi^*}{\gamma N} \gtrsim \frac{1}{\gamma F} \quad (130)$$

So, in the non-binary case we get a higher bound for ρ_* than in the binary case. This is because in the binary case we actually have diagonal $Q_{AB} = N_A \delta_{AB}$ (N_A is defined in Section 5), and $\zeta_A = \sqrt{N_A}$ in that case (for a diagonal factor covariance matrix).

References

- Ackerman, C., McEnally, R. and Ravenscraft, D. (1999) The Performance of Hedge Funds: Risk, Return and Incentives. *Journal of Finance* 54(3): 833-874.
- Agarwal, V. and Naik, N.Y. (2000a) On Taking the “Alternative” Route: The Risks, Rewards, and Performance Persistence of Hedge Funds. *Journal of Alternative Investments* 2(4): 6-23.
- Agarwal, V. and Naik, N.Y. (2000b) Multi-Period Performance Persistence Analysis of Hedge Funds Source. *Journal of Financial and Quantitative Analysis* 35(3): 327-342.
- Amin, G. and Kat, H. (2003) Stocks, Bonds and Hedge Funds: Not a Free Lunch! *Journal of Portfolio Management* 29(4): 113-120.
- Asness, C.S., Krail, R.J. and Liew, J.M. (2001) Do Hedge Funds Hedge? *Journal of Portfolio Management* 28(1): 6-19.
- Brooks, C. and Kat, H.M. (2002) The Statistical Properties of Hedge Fund Index Returns and Their Implications for Investors. *Journal of Alternative Investments* 5(2): 26-44.

Brown, S.J., Goetzmann, W. and Ibbotson, R.G. (1999) Offshore Hedge Funds: Survival and Performance, 1989-1995. *Journal of Business* 72(1): 91-117.

Chan, N., Getmansky, M., Haas, S.M. and Lo, A.W. (2006) Systemic Risk and Hedge Funds. In: Carey, M. and Stulz, R.M. (eds.) *The Risks of Financial Institutions*. University of Chicago Press, Chapter 6, pp. 235-338.

Edwards, F.R. and Caglayan, M.O. (2001) Hedge Fund and Commodity Fund Investments in Bull and Bear Markets. *Journal of Portfolio Management* 27(4): 97-108.

Edwards, F.R. and Liew, J. (1999a) Managed Commodity Funds. *Journal of Futures Markets* 19(4): 377-411.

Edwards, F.R. and Liew, J. (1999b) Hedge Funds versus Managed Futures as Asset Classes. *Journal of Derivatives* 6(4): 45-64.

Fung, W. and Hsieh, D. (1999) A Primer on Hedge Funds. *Journal of Empirical Finance* 6(3): 309-331.

Fung, W. and Hsieh, D. (2000) Performance Characteristics of Hedge Funds and Commodity Funds: Natural vs. Spurious Biases. *Journal of Financial and Quantitative Analysis* 35(3): 291-307.

Fung, W. and Hsieh, D. (2001) The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers. *Review of Financial Studies* 14(2): 313-341.

Kakushadze, Z. and Liew, J.K.-S. (2014) Is It Possible to OD on Alpha? SSRN Working Paper, <http://ssrn.com/abstract=2419415> (April 2, 2014); arXiv:1404.0746.

Kakushadze, Z. (2014a) Spectral Model of Turnover Reduction. SSRN Working Paper, <http://ssrn.com/abstract=2427049> (April 20, 2014); arXiv:1404.5050.

Kakushadze, Z. (2014b) Factor Models for Alpha Streams. SSRN Working Paper, <http://ssrn.com/abstract=2449927> (June 12, 2014); arXiv:1406.3396.

Kakushadze, Z. (2014c) Combining Alpha Streams with Costs. *The Journal of Risk* (forthcoming); SSRN Accepted Paper, <http://ssrn.com/abstract=2438687> (July 7, 2014); arXiv:1405.4716.

Kao, D.-L. (2002) Battle for Alphas: Hedge Funds versus Long-Only Portfolios. *Financial Analysts Journal* 58(2): 16-36.

Liang, B. (1999) On the Performance of Hedge Funds. *Financial Analysts Journal* 55(4): 72-85.

- Liang, B. (2000) Hedge Funds: The Living and the Dead. *Journal of Financial and Quantitative Analysis* 35(3): 309-326.
- Liang, B. (2001) Hedge Fund Performance: 1990-1999. *Financial Analysts Journal* 57(1): 11-18.
- Lo, A.W. (2001) Risk Management For Hedge Funds: Introduction and Overview. *Financial Analysis Journal* 57(6): 16-33.
- Rebonato, R. and Jäckel, P. (1999) The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. SSRN Working Paper, <http://ssrn.com/abstract=1969689> (December 7, 2011).
- Schneeweis, T., Spurgin, R. and McCarthy, D. (1996) Survivor Bias in Commodity Trading Advisor Performance. *Journal of Futures Markets* 16(7): 757-772.
- 't Hooft, G. (1974) A Planar Diagram Theory For Strong Interactions. *Nuclear Physics* B72(3):461-473.

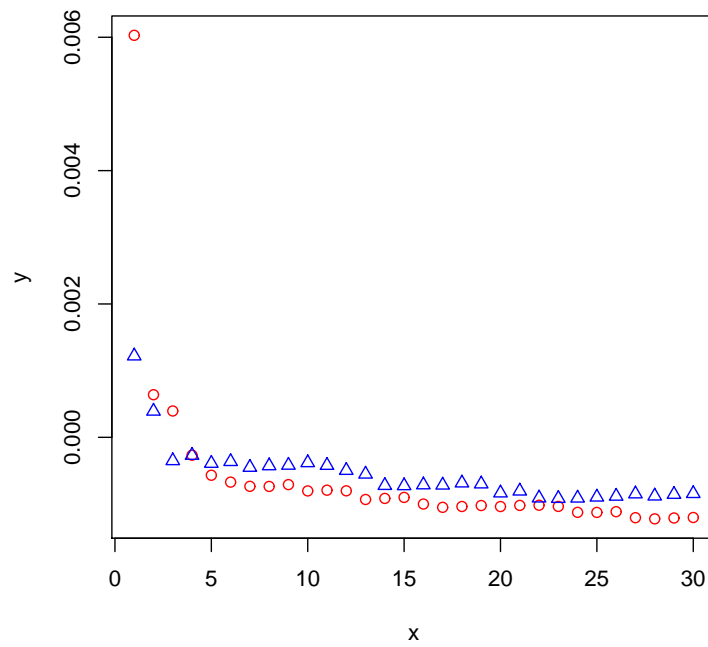


Figure 1: x -axis: K ; y -axis: ζ_1 . Circles: raw HF; triangles: HF regressed over Mkt-RF and Fama-French risk factors SMB, HML, WML. See Section 7 for details.

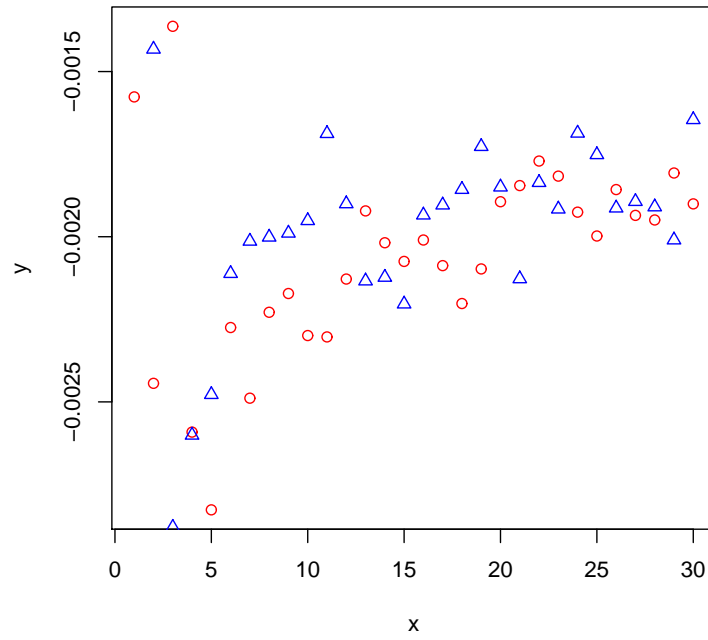


Figure 2: x -axis: K ; y -axis: ζ_2 . Circles: raw HF; triangles: HF regressed over Mkt-RF and Fama-French risk factors SMB, HML, WML. See Section 7 for details.

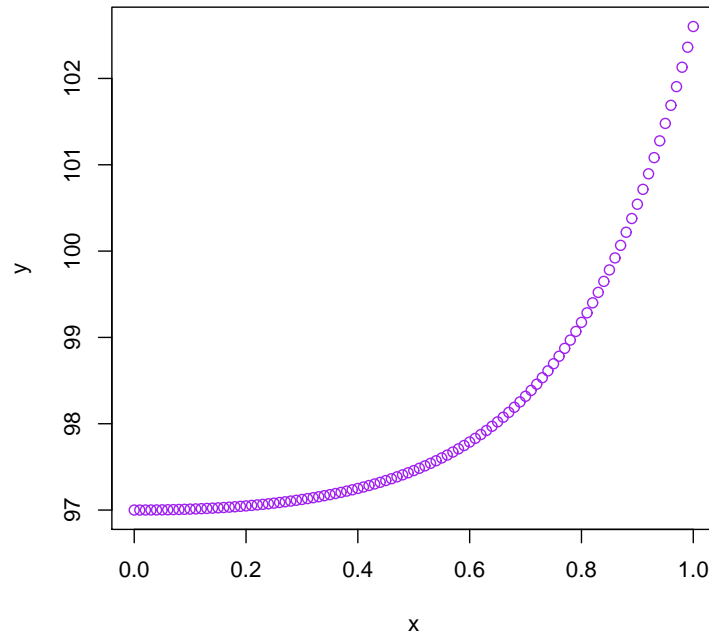


Figure 3: x -axis: ρ ; y -axis: $\hat{\psi}^*$. See Appendix A for details.